

## What Makes Weak-Willed Action Weak?

Tamar Schapiro, MIT

[Forthcoming in *Normativity and Agency: themes from the philosophy of Christine M. Korsgaard*, edited by Tamar Schapiro, Kyla Ebels-Duggan, and Sharon Street (Oxford, 2022). Please do not cite these page numbers – this is not quite the final version.]

**Abstract:** Most accounts of weak-willed action characterize it as an error in deliberation. The agent acts on the weaker rather than the stronger reason, or the apparent rather than the actual reason, or the partial rather than the all-things-considered ought. That might capture the idea that weak-willed action is bad or wrong – perhaps even bad or wrong, by the agent’s own lights. But in what sense is weak-willed action weak? Weakness and strength are notions that we define in relation to some sort of force or pressure. The weak-willed agent is one who gives in to this pressure. But what, in principle, could pressure a free will? I argue that the only thing that can pressure a free will is the burden of freedom itself. To act in a weak-willed way is to flee the burden of one’s freedom. Weakness of will is not a mistake within deliberation, but rather a flight from it.

<1> I. The question

Weak-willed action is often characterized as action against one’s own better judgment. You know you shouldn’t drink another cocktail, but you drink it anyway. You know you shouldn’t put off fixing your car, but you put it off anyway. You know you shouldn’t spend so much money, but you spend it anyway. Acting in this way amounts to a kind of failure, by your own lights. You are doing something wrong or bad or, at least, deficient, according to your own standards. But the characterization of these actions as “weak” suggests something further. It suggests that in violating your own standards, you are also giving in to some kind of pressure. Strength and weakness are notions that we define in relation to some kind of force. We use muscle power to oppose the force of gravity. What sort of force does will power oppose? What does the strength and weakness of a will consist in?<sup>1</sup>

The standard characterization of weak-willed action, as action against one's better judgment, tends to make it hard to see this question clearly. Consider the example I just gave. You know you shouldn't drink another cocktail, but you drink it anyway. Notice that in picturing this scenario, we tacitly assume the situation is one in which you *want* to drink another cocktail, in the sense of *feeling like it*, or *being inclined* to do so. Suppose it were otherwise. You know you shouldn't drink another cocktail, *and you have no inclination to do so*, but you drink it anyway. How odd. Described in this way, the scenario is hardly a paradigmatic case of weak-willed action. It is paradigmatic if we assume that in acting against your better judgment, you are giving in to your inclination. And yet philosophers standardly leave this last clause out of the description of the explanandum. This makes it hard to even raise the question I am interested in. What sort of pressure are you giving in to, when you go along with your inclination, against your better judgment? What sort of counterforce is required to oppose it?

In characterizing such action as weak, it is important to avoid embracing a picture according to which one is literally overpowered by a physical force. I am assuming that when you drink another cocktail, against your better judgment, your behavior is not simply compelled. It is not the equivalent of a hiccup or a seizure. You are not physically forced to bring the glass to your lips, against your will. Granted, there may well be such thing as responsibility undermining dependence on substances like alcohol, nicotine, and opioids, as well as responsibility-undermining compulsive conditions, like Tourette's syndrome. Just how to conceive of these conditions is a separate philosophical question. My interest in weak-willed action is an interest in an ordinary, nonpathological type of failure, for which we rightly, and

routinely, hold ourselves and others responsible. My point here is that in thinking about this ordinary type of failure, it is unhelpful to conceive of ourselves as being literally overpowered by

184

a brute force. That picture may seem to capture the sense in which weak-willed action is weak, but it does so by sacrificing the sense in which such action is willed.

The standard alternative is to characterize weak-willed action as a type of deliberative error—a violation of one’s own standards of practical thinking. The accounts I have in mind have the following structure: they specify a conception of proper deliberation, and then identify the specific way in which the weak-willed agent diverges from this procedure. According to Aristotle, the weak-willed agent fails to properly grasp the practical syllogism.<sup>2</sup> On Donald Davidson’s account, the agent violates the principle of continence, which tells her to act on her all-things-considered judgment.<sup>3</sup> More generically, the agent is characterized as acting on the weaker rather than the stronger reason, or on the partial rather than the comprehensive reason, or on the apparent rather than the actual reason.

All of these accounts avoid describing the agent as being overpowered by a brute physical force. Instead they preserve the sense in which weak-willed action is willed, by characterizing the agent as actively engaged in practical thinking. But then in what sense is such action weak?

Granted, the agent is violating her own deliberative standard, and in that sense she is acting against her own better judgment. But in making this mistake, how is she giving in?

<1> II. An acute form of the worry

To see this worry in its most acute form, it is helpful to understand how it arises within Kant’s

theory of action. A central feature of Kant's theory of motivation is his insistence that the will is free, in the sense that it is always in principle possible for us to choose to refrain from

185

satisfying our inclinations.<sup>4</sup> What is sometimes overlooked is a further aspect of this freedom, commonly referred to as Kant's "Incorporation Thesis."<sup>5</sup> Kant writes:

...freedom of the power of choice has the characteristic...that it cannot be determined to action through any incentive *except so far as the human being has incorporated it into his maxim* (has made it into a universal rule for himself...); only in this way can an incentive, whatever it may be, coexist with the absolute spontaneity of the power of choice (of freedom).<sup>6</sup>

Without getting too far into the terminological weeds, the basic idea is that insofar as you have an inclination to  $\varphi$  (insofar as you "feel like  $\varphi$ -ing"), you are not yet  $\varphi$ -ing.<sup>7</sup> Your inclination, simply as such, *cannot* determine you to act. If you are to act on your inclination, you must determine yourself to do so. You must freely accept, or endorse, or choose to act on your inclination. In Kant's terminology, you must "incorporate" your "incentive" into your "maxim." Later I will say more about how I think these specific terms should be interpreted, but for now the basic idea is what matters. Kantian freedom is not just the freedom to choose to refrain from acting on the inclinations you happen to have. It is also the burden of not being able to simply allow your inclinations to do the choosing for you.

The Incorporation Thesis implies that weak-willed action always involves some sort of free choice. That makes it particularly challenging to explain how weak-willed action is weak. Like the standard theorists I discussed earlier, Kant seems to characterize weak-willed action as a kind

of deliberative mistake. While there is much room for controversy about exactly what the mistake is, it involves some sort of failure to give the moral law priority over self-love.<sup>8</sup> Kant holds that all human agents are necessarily committed to these two basic practical principles. Proper deliberation involves making the moral law the unconditioned condition of self-love, in

186

the sense that you commit to satisfying your inclinations only on the condition that you can do so without violating the moral law. The weak-willed agent at some level reverses this ordering. She satisfies her inclination regardless of whether doing so is consistent with the moral law, and so gives self-love priority. This is a mistake by her own lights, because as Kant sees it, she is at the same time necessarily—if only implicitly—committed to the supremacy of the moral law. Suppose we can give a coherent account of this deliberative error. Such an account might preserve the sense in which weak-willed action is willed. But can it explain how weak-willed action is weak? If the agent is free in the sense spelled out by Kant's Incorporation Thesis, then even a strong inclination cannot, in principle, determine her to reverse the proper order of the two principles. And if that is so, then what could it mean to say, even in principle, that her inclination puts pressure on her will?

*What, in principle, could pressure a free will?*

### <1> III. Rational pressure

Kant himself does not give us a developed theory of the relation between inclination and will.<sup>9</sup>

Indeed he even seems to talk about the influence of inclination in various incompatible ways.

Sometimes he uses a gravitational metaphor. He describes our inclinations as providing the

“counterweight” to the commands of duty.<sup>10</sup> And he writes that because human nature is affected by inclinations, “virtue can never settle down in peace and quiet with its maxims adopted once and for all but, if it is not rising, is unavoidably sinking.”<sup>11</sup> The gravitational metaphor is clearly in tension with Kant’s Incorporation Thesis, because we can in principle be helpless to overpower gravity, while we cannot in principle be helpless to refrain from acting on

187

the inclinations we have. That said, Kant does not rely solely on the gravitational metaphor. Sometimes he describes inclinations as “claims” that one part of ourselves addresses to the rest of ourselves.<sup>12</sup> Could this idea help us to understand weak-willed action?

Suppose that, in being inclined to drink another cocktail, you aren’t just being pushed around. Your inclination does not influence you as a brute force. Instead it operates as a kind of lens through which you “see” certain considerations as reasons. When you are inclined to drink another cocktail, you are, let’s say, “seeing” the pleasant effects of drinking another cocktail as a reason to drink one.<sup>13</sup> The idea is that the motivational pressure of your inclination is not the pressure of a brute, physical force, but rather the normative pressure of a reason, as seen through the lens of your inclination.

What is the normative pressure of a reason? Consider the pressure you feel when you understand the premises of a valid argument, but you have yet to draw the conclusion: All human beings are mortal.

Your beloved is a human being.

Insofar as you recognize the soundness of the premises, your mind is subject to pressure to draw the conclusion: your beloved is mortal. Call this rational pressure. Perhaps the form of pressure

in relation to which the will can be strong or weak is rational pressure, in this sense. Does that help us understand the sense in which weak-willed action is weak? The idea would be that in drinking another cocktail, you are seeing the situation as warranting this action. Your awareness of salient features of the situation is analogous to your recognition of the soundness of the premises, and your action is like your drawing of the conclusion. In giving in, you are giving in to rational pressure.

188

Does this picture stand up to scrutiny? We can imagine it in two different ways. In the first scenario, you are not aware of the insufficiency of the reason you are acting on. You are aware of the consideration that drinking another cocktail would be pleasant, and you are taking that consideration as a reason to drink it, but you are not aware of any stronger reason to refrain. In that case, you are not making a mistake, by your own lights. You are just acting on the reasons, as you see them. You might be wrong, but you are not violating your own deliberative standard. You are acting on a reason that you take to be sufficient. If there is any sense in which you are “giving in,” you are giving in to the pressure of the strongest reason you are aware of. That is hardly a notion of weakness, in the pejorative sense.

Alternatively, we can imagine that you are aware of the insufficiency of this reason, even as you act on it. You are acting on what you take to be the weaker reason. If that is what is going on, then you are making a mistake by your own lights. But in doing so, what are you giving in to? The answer cannot be that you are giving in to the normative pressure of a reason, as you see it. The stronger reason, by your own lights, supports forgoing the cocktail. In acting on the weaker reason, you must be giving in to some other kind of pressure.

One could argue, perhaps, that in giving in to the weaker reason, you are giving in to something analogous to an optical illusion. Your inclination is not just a lens, but a distorting influence, through which weaker reasons appear to you as if they were stronger, just as a straight stick in the water appears bent. When you are inclined, against your better judgment, to drink another cocktail, you are acting on the illusion that you have stronger reason to drink.<sup>14</sup> Does this appeal to perceptual illusion help?

189

Again, we can imagine the scenario in two different ways. Either you are unaware of the situation as involving any distorting influence, or you are aware of it as such. In the first version, you are seeing the situation as rationally warranting you drinking another cocktail, and you have no awareness of any distorting influence at work in this perception. If that is the situation, then in drinking another cocktail, you are not making a mistake by your own lights. And you are not giving in to any force other than the force of the stronger reason as you see it. There is nothing weak about that.

On the other interpretation, you are aware of the illusion. You are “seeing” the situation as giving you reason to drink another cocktail, but you are at the same time aware of this perception as the product of a distorting influence. You are aware of it as *mere* appearance. If you nevertheless give in to the force of this appearance, what are you giving in to? I think here the role of the distorting influence is analogous to the role of a brute force. Either it simply overpowers you, in which case you are simply brainwashed. Or you are aware of it as a feature of your circumstances, something you are free to choose to at least try to resist. Just as you are not forced to conclude, on the basis of what you regard as *mere* appearance, that the stick in the

water is bent, so you are not forced to conclude, on the basis of what you regard as *mere* appearance, that you have stronger reason to drink another cocktail. You can at least try to resist. And by hypothesis you take yourself to have reason to do so. That reason is, by your own lights, stronger than any reason you might have to go along with an influence that you yourself regard as distorting. If you nevertheless give in, then either you were overpowered, or you acted on the weaker reason. And we still do not have an account of what you are giving in to when you do that.

<1> IV. Having an inclination

If the relevant form of pressure is neither brute force nor rational pressure, then what is it? I am going to suggest an answer that I think Kant would embrace. *The only thing that can pressure a free will is the burden of freedom itself.* Inclinations, as such, cannot put pressure on us, in our role as free choosers of our actions. The experience of frustrating an inclination is, no doubt, painful. But as free choosers, it is up to us to decide which pains are worth enduring. A given inclination, simply as such, cannot give us reason to judge that the pain of frustrating it is, or is not, worth enduring. But inclinations, simply as such, do provide us with the opportunity to flee the burden of our freedom. Let me explain.

I will start by giving an account of what it is to have an inclination, prior to determining yourself to act on it. This is something I call “the moment of drama,” because in this moment, you are free to act on your inclination or not. You are in the moment of drama when you are having an inclination to drink another cocktail, but you have not yet determined yourself to do

so. In this moment, you are free.

The sense of freedom I am invoking here is not the contra-causal freedom imagined by the incompatibilist. It is the freedom to choose among the actions that depend on you, the actions you could do if you chose to do them. A full defense of this conception of freedom would show that it is not threatened by causal determinism, or by mechanism more broadly. It would also show that attributing this sense of freedom is not arbitrary. It is both necessary and legitimate, but only insofar as we have to engage in practical reasoning.<sup>15</sup> I will not try to offer such a defense here. The point for my purposes is that freedom in this sense is something we have to attribute to ourselves insofar as we find ourselves faced with the task of deciding what to do.

191

Insofar as you find yourself having an inclination to drink another cocktail, you are faced with that task.<sup>16</sup>

To understand just what this task involves, we need a clearer picture of what you are faced with when you are inclined, but not yet determined, to do something. We need a clearer account of what I am calling the moment of drama, or the moment of “having an inclination” to do something. On the theory of inclination that I have developed elsewhere, to have an inclination to  $\varphi$  is “to be drawn out of yourself.”<sup>17</sup> A part of you is seeing and responding to the world in a distinctive way, while the rest of you is aware of that activity. You are not in conflict with yourself, but you are aware of a division within yourself, such that there is something you call an “inclination” that you are “having.” To illustrate: when you are having an inclination to drink another cocktail, prior to determining yourself to act, part of your conscious mind is spontaneously focusing on the cocktail, real or imagined, in a distinctive way (I will say more about this in a moment). Moreover, this part of your mind exerts a guiding influence over your

nonvoluntary agential capacities. You may start salivating, and some of your muscles may contract, as if initiating the pursuit of the cocktail. These conscious and physiological responses, though nonvoluntary, are purposively structured. They are oriented towards drinking another cocktail. In this respect, they are unlike sneezes and hiccups. When you are aware of yourself having an inclination, I claim, you are aware of a form of activity of that is purposively guided by a “someone” within, a part of your mind that is not your free, choosing self.<sup>18</sup>

The reason I am claiming that this “someone” is not your free, choosing self, is that your inclination is not attributable to your will. You can choose to drink another cocktail, and thereby drink another cocktail. But you cannot choose to feel like drinking another cocktail, and thereby feel like drinking another cocktail. What we choose to do is up to us. What we feel like doing is

192

not up to us, in the same sense. This is a fundamental feature of our human predicament as agents. We can strive to cultivate our inclinations over time, and with luck, we can end up feeling like doing the things we choose to do. But we cannot simply feel like doing things at will. We “have” inclinations that are not the same as our actions because, for better and for worse, part of us has a practical mind of its own.

This subagential part of us can initiate activity independent of our choosing selves. But it does not have a life of its own to lead. In this respect, it is not analogous to a separate individual who might happen to be located within the boundaries of your natural organism. Rather, it is more closely analogous to an organic system, like your digestive system. Your digestive system can initiate activity independent of your choosing self. But it does not have a life of its own to lead. Its function is to contribute to your life. Now, your inclinations are consciously guided, in a way that your digestive activity is not. (I will say more about this shortly.) This is why it makes

more sense to think of the source of your inclinations as a kind of “someone.” But that someone does not have any claim to a life of its own. Its function is to contribute to your life. I call this someone your “inner animal,” because I think it makes sense to understand it as having the structure of what Christine Korsgaard calls an “instinctive mind.”<sup>19</sup> As Korsgaard characterizes it, the mind of a creature of instinct relates itself to a “teleological” world, a world that is infused with purposiveness. More specifically, this mind sees a world that is infused with its own purposiveness. It is a world structured in terms of that creature’s purposes. In Korsgaard’s words, the instinctive mind sees objects as “to-be-fled,” or “to-be-eaten.” In this respect, she claims, this mind differs from the “rational mind,” which does not see the world teleologically. Exactly how to fill this out requires interpretation, but what Korsgaard is getting at is the idea that the instinctive mind neither can, nor has to, raise a certain kind of question

193

about how it should respond to the world. Its responses are built into its nature, and are projected onto its way of seeing its world. The rational mind, by contrast, does not have its responses built into its nature in the same sense, nor does it project its responses onto the world in the same sense. Because of this, the rational mind is forced to ask a certain kind of question that the instinctive mind cannot ask, namely, a question about how it *should* respond to the world.

To dig a bit further into the details: as I mentioned, Korsgaard characterizes the instinctive mind as seeing objects as “to-be-fled,” “to-be-eaten,” etc. But I believe what Korsgaard intends, and needs for her argument, is a characterization that is even a bit more alien than this. In saying that the animal sees an object as “to-be-fled,” she risks suggesting that the animal has the capacity to separate the object, conceived non-teleologically, from the property, conceived teleologically. She risks suggesting that the antelope can see, for example, that

approaching lion as simply a thing (viewed non-teleologically), to which the (teleological) property of to-be-fledness happens to attach. And this in turn might suggest that the antelope could, at least in principle, raise the question, “is this lion (viewed non-teleologically) really to be-fled?” To fully appreciate the idea that the nonhuman animal sees the world teleologically, I think it is better to characterize the animal as seeing, for example, “lion-y-to-be-fledness.” The structure of the thought is not that of non-teleological object plus teleological predicate. Rather, the package as a whole is thoroughly teleological.<sup>20</sup>

Turning back now to the theory of inclination, my claim is that when you are inclined, say, to drink another cocktail, part of your mind is seeing and responding to another cocktail in this instinctive mode. Your nonvoluntary agential capacities, conscious and physiological, are purposively guided by a thought of the form, “cocktail-y-to-be-drunkenness.” The effects of this activity constitute “what it feels like” to be inclined to drink another cocktail. Your inner animal

194

is active, while the rest of you, the one who is “having” the inclination, is as yet undetermined. You are, in this sense, drawn out of yourself.

This is not yet an account of weak-willed action. It is an account of what it is to have an inclination to do something. I have not assumed in this description that your inclination runs contrary to your better judgment, nor have I said much about what it takes to go from having the inclination to acting on it. But just this much of the theory might suggest a familiar picture of the human agent, according to which reason is the rider, and passion is the horse.<sup>21</sup> Is that the model?

That model would be problematic, given what I have already argued. The rider/horse picture implies that our inclinations influence us externally, as brute forces that wholly bypass the will. So what picture is more accurate? I think it is a significant philosophical challenge to

lay out a conception of the relation between your inclinations and the part of you that “has” them. It is tempting to try to do this by assimilating that relation to something more familiar, but ultimately, I think our relation to our inclinations is *sui generis*. Analogies to more familiar relations cannot capture what is distinctive about being a rational animal.<sup>22</sup> Still, I think I can say more about what it takes to go from having an inclination to acting on it, and what I say about that will help us answer the question I have been raising about what makes weak-willed action weak.

#### <1> V. Incorporation

Given the inner animal view as I have laid it out thus far, how are we to conceive of the task we face when we are in the moment of drama? When you find yourself having an inclination to drink another cocktail, your inner animal is guided by its instinctive awareness of

195

“cocktail-y-to-be-drunkenness.” I am going to use the term “incentive” to refer to this thought. Your incentive is the instinctive thought that is guiding your inner animal when you are having an inclination. Now recall that Kant’s Incorporation Thesis holds that your incentive, simply as such, cannot determine your will. Why not?

Kant does not spell out why not. This is because he does not say clearly enough what an incentive is, or how it is related to your will. Now, I have just given a theory of what an incentive is, at least insofar as we have in mind an incentive of inclination.<sup>23</sup> An incentive of inclination is the instinctive thought that guides the activity that constitutes your inclination, such as “cocktail y-to-be-drunkenness.” I want to claim that although “cocktail-y-to-be-drunkenness” is fit to

guide instinctive thinking, it is not fit to guide the thinking of your deciding self. This is because there is a mismatch between the form of the incentive, and what you need to function as a free agent. As a free agent, you have to choose your actions. To choose your actions, you have to first represent them to yourself, as possible objects of choice. My claim is that your inner animal's thought, "cocktail-y-to-be-drunkenness," is not a representation of a possible action. It is not fit to play the role of a possible object of practical choice. To see this, consider the fact that it makes no sense to ask, "should I cocktail-y-to-be-drunkenness?" "Cocktail-y-to-be-drunkenness" does not describe a possible action as a kind of thing—an object of reflection, something that can be looked at, critically evaluated, and chosen or rejected. In this sense, there is a mismatch between the form instinctive thinking takes, and the form of thinking required to engage in practical deliberation.<sup>24</sup>

I believe this is why something Kant calls "incorporation" is necessary. When you are having an inclination to drink another cocktail, your incentive is "cocktail-y-to-be-drunkenness." If you are to make the transition from having that inclination to acting on it, you have to

196

transform "cocktail-y-to-be-drunkenness" into something that you can deliberate about. You have to make it into a representation that can serve as an object of your choice. I believe this is the role of what Kant calls your "maxim." Your maxim, as I understand it, is a description of your action that can serve as a possible object of your choice. "Cocktail-y-to-be-drunkenness" cannot play this role, but something like "drink another cocktail" looks like it can. It at least makes grammatical sense to ask, "should I drink another cocktail?" So as a first pass, we can say that in order to go from having an inclination to acting on it, you have to transform your incentive, "cocktail-y-to-be-drunkenness," into a maxim, where that is something like, "drink

another cocktail.”

Now, this isn't quite right, because “drink another cocktail” is not yet a maxim in Kant's sense. A maxim, in Kant's sense, includes something about the agent's reasons for choosing to do what she does. It is a description of your action, under which you find it worthy of your choice. So construed, the notion of a maxim is the correlate of Korsgaard's notion of a “practical identity.”<sup>25</sup> A practical identity is a description of who you are, under which you value yourself. A maxim, I want to say, is a description of what you are doing, under which you value it. And just as your practical identity sets the standard of your integrity as a person, your maxim sets the standard of the integrity of your action. It is with reference to your maxim that you judge whether or not you are successful in doing what you have chosen to do.

If that is the right way to understand the notion of a maxim, then “drink another cocktail” is not yet fit to be one. “Drink another cocktail” is simply an outward description of an action. It says nothing about what you might find choiceworthy in drinking another cocktail. So if you are to incorporate your incentive into your maxim, you have to do more than transform “cocktail-y to-be-drunkenness” into an outward description, like “drink another cocktail.” You have to

197

transform it into something like, “drink another cocktail, in this way, at this time, for this reason, etc.” You have to represent your action to yourself as the specific form of cocktail-drinking that you could find choiceworthy. Your maxim is that more specific description.

Let me construct a scenario to make this more concrete. Suppose the context in which you are inclined to drink another cocktail is your child's wedding reception. Even though you know you will likely incur a headache in the morning, you consider that it might be worth it to have another drink. You have been very anxious in anticipation of this event, worrying about all

the logistics, and now you just want to be able to relax a bit and enjoy each moment. In considering this choice, you are representing the possible action to yourself in a certain way. You are thinking about drinking another cocktail, given that this is a special occasion, for the purpose of celebrating the moment more deeply, on the condition that you are not likely to become objectionably inebriated, in a way that might embarrass yourself or anyone else, and on the condition that you are not going to get behind the wheel of a car. That is your maxim.<sup>26</sup> It is much more detailed than a generic outward description of the action. Indeed, it describes the specific version of cocktail-drinking that you can claim as the object of your possible choice in this situation.

To summarize, then, on my interpretation of Kant's Incorporation Thesis, your incentive, "cocktail-y-to-be-drunkenness," is unfit to serve as the object of your choice, because it is not a representation of an action. This tells us something about the structure of the task you face, when you are inclined, but not yet determined, to drink another cocktail. In order to act on your incentive, you have to make it into a representation of an action that could be the object of your choice. You can do this by taking the instinctive thought, "cocktail-y-to-be-drunkenness" as motivational raw material out of which to construct your maxim. I call this "humanizing" your

198

incentive. You are transforming a guiding thought that is fit for an instinctive mind into a guiding thought that is fit for the mind of a free agent. To "incorporate" your incentive into your maxim is to humanize it, in this sense.

<1> VI. Giving in

I have claimed that when you are having an inclination to do something, you are faced with the following task: can you transform your incentive into a representation of a possible object of your free choice? This will not always be possible. Suppose you are inclined to drink another cocktail at your child's wedding, but you are committed to driving family members home at the end of the evening. In that case, there may be no version of drinking another cocktail that you can choose, consistent with your values. You are inclined to drink another cocktail, but given your commitment to being a conscientious driver, you simply have to forgo the cocktail. Fair enough. But notice what Kant's Incorporation Thesis seems to imply in that situation. It seems to imply not only that you *ought* not drink another cocktail, but that you *cannot*. Kant's claim is that your incentive *cannot* determine your will unless you incorporate it into your maxim. If it is impossible to incorporate your incentive into your maxim, then presumably you cannot act on that incentive. But of course you *can* act on it. You can certainly drink another cocktail, against your better judgment, even though you have committed to drive family members home. What are you doing when you are doing that?

I think Kant should have said more here, but he didn't.<sup>27</sup> I want to show that the inner animal view provides an answer. When you are having an inclination to drink another cocktail, you are, as Kant says, "at a crossroads."<sup>28</sup> I want to characterize the crossroads in this way: you

can accept the burden of your freedom, or you can flee that burden. If you accept it, you take "cocktail-y-to-be-drunkenness" as motivational raw material out of which to try to construct your maxims. That is, you take your incentive as a kind of resource, out of which to actively construct

your conduct. You take it upon yourself to humanize your impulse as a condition of acting on it. In this sense, you affirm your position as governor of yourself, leader of your life. But there is an alternative path open to you.<sup>29</sup> You can flee the burden of your freedom, by fleeing into your animality. Instead of taking it upon yourself to humanize your impulse, you dehumanize yourself. How might this be possible?

You cannot choose to become a nonhuman animal, and thereby become a nonhuman animal. You can fall into a condition that is something like that of a nonhuman animal, either due to injury or disease, or as a response to a situation that your organism registers as an emergency. But you cannot put yourself into this condition at will. What I am trying to describe in weak willed action is a way of freely putting yourself into the position of a nonhuman animal, while all the while not being one.

When you take this low road, you inhabit the instinctive part of your mind, as if it were your whole self, in such a way as to avoid taking responsibility for what you do. Your inner animal is already guided by cocktail-y-to-be-drunkenness, and you, the one “having” the inclination, allow yourself to be so guided. You act as if the instinctive part of your mind were your whole mind. You act as if it were simply your nature *to be a drinker of cocktail-y-to-be drunkenness*. In this sense, you act as if you have no choice in the matter of what to do with your incentive.

To see how this way of inhabiting your inner animal’s mind counts as fleeing your freedom, I want to contrast it with a different way of inhabiting your inner animal’s mind. The

former I will call “giving in,” while the latter I’ll call “loosening up.” There is nothing wrong with loosening up. Suppose you are dancing at your child’s wedding. You find yourself

becoming self-conscious, thinking too much, and failing to enjoy the moment. So you decide to try to loosen up. You try focusing more directly on the music, tuning out all the other people in the room, so as to allow your body to move more spontaneously. As you are doing this, you are aware of limitations. You are not going to move in such a way as to step on someone else's toes, or make a complete spectacle of yourself. But you do make an effort to quiet your deciding mind, so as to allow your instinctive mind to have greater latitude in guiding you. With any luck, you will stop overthinking and start "going with the flow."

In loosening up, you are trying to give your instinctive mind a greater role to play in shaping what you do. But notice that this is entirely consistent with taking responsibility for what you do. You are choosing to loosen up, as a good way of doing something you value, namely enjoying yourself at your child's wedding. Loosening up is thus something you do at the executive, rather than the legislative stage of action. It is something you do, as a way of more effectively doing what you have chosen to do.

Giving in is different. It is not something you do at the executive stage, as a way of more effectively doing what you have chosen to do. It is, rather, something you do instead of choosing to do anything. When you give in, you are not enlisting your inner animal as a resource, something you can use to enhance your effectiveness in leading your life. Instead you are abandoning your post as the leader of your life, by escaping into your instinctive mind. Your inclination does not force you to do this. But it creates the opportunity for you to do this, by making you aware of an instinctive mind to flee into.

you are choosing to avoid the pain of frustrating your inclination? The idea that it is painful to thwart an inclination, simply as such, makes sense on my view. Recall that in having an inclination, your non-voluntary capacities are already purposively engaged, and you experience this physiologically and consciously. Once non-voluntary activity is underway, there is a sense in which thwarting that activity before it has reached completion is painful. So why not just say that in giving in to your inclination, you are choosing to avoid this pain?

As a free agent, it is up to you to decide which pains are worth avoiding, and which pleasures are worth pursuing. Sometimes it can make sense to act on an inclination, in part because enduring the pain of resisting it is simply not worth it to you. When that is the case, your maxim will reflect this consideration. But when you simply give in to your inclination, you are acting as if it is your nature to be an avoider of pain, simply as such. In this sense, you are fleeing your freedom.

When you give in, are you acting without a maxim? On my view, the answer is that you are acting on a defective maxim. You are taking the instinctive thought, “cocktail-y-to-be drunkenness,” and forcing it into the role of your maxim, even though it is unfit to play this role.

As a consequence, your action is defective, *qua* action. What I mean is not that you are not responsible for it. It is not like a hiccup or a sneeze. But there is a sense in which you are acting without really choosing your action, and without really knowing what you are doing. As I have emphasized, “cocktail-y-to-be-drunkenness” is not a description of an action, and it cannot really serve as an object of choice. But insofar as you can inhabit your instinctive mind, pushing aside the awareness of your capacity for free decision, you can still press this raw incentive into the

role of a maxim. When you do this, you are coming as close as is humanly possible to acting without choosing, and to acting without knowing what you are doing. But since you are not externally coerced or deceived—since the defect in your action is your own doing—you are fully responsible.

I just said that when you give in, you push aside the awareness of your capacity for free decision. What I mean to note is that in giving in, you are necessarily engaged in a kind of self-deception.<sup>30</sup> You are not simply deceiving yourself about which consideration is the stronger reason.<sup>31</sup> To put it this way is to make weak-willed action sound like a mistake within deliberation. On my view, weak-willed action is a flight from deliberation. To the extent that you are deceiving yourself, you are deceiving yourself about your nature. You are acting as if you are not burdened with the freedom of having to construct and choose your actions.

Rationalization, in the pejorative sense, is made possible by this more fundamental form of self-deception. What is it to give a rationalization of your weak-willed action? I want to resist the idea that a rationalization is just an inaccurate account of the deliberation that led you to act as you did. When you offer a rationalization, to yourself or to others, you are not just being insincere about which reasons you acted on. More fundamentally, you are being insincere about having acted on reasons at all. Instead of deliberating, you simply gave in, and then you made up a story about having had reasons for what you did.

This may make it sound as if, on my view, weak-willed action is something essentially impulsive, something that cannot be calculated and instrumentally sophisticated. That is actually not implied by my view. Suppose you give in to your inclination to drink another cocktail. If there are no cocktails close to hand, you will have to find one. To do this, you might have to strategically make your way through a crowd of wedding guests, politely disengaging from any

lengthy conversations, and targeting the bartender who can serve you most efficiently. To this extent, you have to know how to master objects in your environment. You have to approach those objects as resources for you to use, and you have to use them effectively. But mastery over your environment should not be confused with mastery over yourself. You can be fleeing your freedom internally, while making effective technical use of the means at your disposal.<sup>32</sup>

#### <1> VIII. Another kind of weakness: going along

Inclinations do not pressure you, *qua* free will. Instead they give you the opportunity to flee into your animality. To do so is to flee into something Sartre called your “facticity.” It is to act as if your nature is something fixed and given, something that determines your choices. Fleeing into your animality is one way of fleeing into your facticity. But it is not the only way. Another way to dehumanize yourself is to flee into automaticity.

Our social environment provides us with scripts to follow. Our local cultures and subcultures are shot through with norms that tell us how to interpret the situations we are in, and how to respond to those situations, so interpreted. They tell us what to say and what to do, independent of what we may happen to feel like doing. Once we have internalized these norms and scripts, they exert a motivational pull on us, a pull that can, in principle, be at odds with what we happen to feel like doing. My claim is that these socially scripted incentives, like incentives of inclination, put us at a crossroads. Insofar as we accept our freedom, we regard these scripts as raw material out of which to construct our actions. We choose to make specific features of our cultures and subcultures our own, and perhaps to reject others, according to our values. But when

flee our freedom, we act as if we were simply built to do and say what we are expected to do and say. We allow ourselves to act as if we were cogs in the social machine.

Let me explain in more detail. Socially scripted incentives are not products of an instinctive mind. They do not have the form, “cocktail-y-to-be-drunkness.” Rather, they generally specify actions as outward performances. “Smile!” “Act like a lady!” “Dress professionally!” But as I argued earlier, to describe an action as an outward performance is not yet to construct a maxim. When you take the high road with respect to an incentive like “act like a lady!” you make a decision about, say, whether and in what respects you value acting like a lady. If you decide to act like a lady, you decide to do so only in this way, for this purpose, under these conditions, given these qualifications, etc. In doing so, you find a way of acting like a lady that counts as your own. Or, if there is no version of acting like a lady that you can value, you dispense with that incentive entirely, and find some other incentive to use as your starting point. But when you take the low road, you flee into automaticity. You act as if it is simply your function to act like a lady, and as if you have no choice in whether and how to do so. This is a kind of weak-willed action, even if it does not fit the paradigm of indulging your inclinations against your better judgment. It is the kind of weakness that Nietzsche and Emerson were centrally concerned about, the weakness that consists in blindly following the herd or the crowd. And it is arguably the kind of weakness that Hannah Arendt and Stanley Milgram were concerned about as they tried to make sense of human nature in the wake of the Holocaust. The forms of evil they describe are not mistakes within deliberation. They are flights from deliberation, and from our shared humanity.

This point is actually broader than it appears. Customs, habits, and routines are sources of scripts that can function in the same way. Those can be shared or idiosyncratic. Donald Davidson

205

describes someone who has no inclination to get up and brush his teeth, but who does so anyway, against his better judgment (because what he really needs at the moment is sleep).<sup>33</sup> Davidson uses this example to argue that weakness of will need not involve giving in to one's inclinations. I agree. As I read it, this person has allowed himself to flee into automaticity. He has allowed himself to become a cog in a machine of his own making. Routines can be helpful when employed at the executive stage of action. They can allow us to "go on autopilot," for the sake of being more effective at doing what we have chosen to do. But they can also provide us with the opportunity to avoid deciding what to do in the first place. Autopilot is wonderful as a tool for an autonomous person to use. It is not a substitute for autonomy.

## <1> VIII. Conclusion

One aim of this paper has been to highlight a constraint that any theory of weak-willed action has to meet. It has to show not only the sense in which weak-willed action is willed, but also the sense in which it is weak. Most theories that do the former characterize weak-willed action as a kind of mistake within deliberation. But they fail to show any sense in which the agent making this mistake is weak. My suggestion is that weak-willed action is weak because it is a flight from freedom. Whether we are fleeing into our animality ("giving in to feel good") or into automaticity ("going along to get along") what we are doing, more fundamentally, is fleeing the burden of having to lead human lives.

206  
Bibliography

- Allison, Henry. *Kant's Theory of Freedom*. Cambridge: Cambridge University Press, 1990.
- Baron, Marcia. "Freedom, Frailty, and Impurity." *Inquiry* 36, no. 4 (1993): 431-441. Bok, Hilary. *Freedom and Responsibility*. Princeton, NJ: Princeton University Press, 1998.
- Davidson, Donald. "How is Weakness of the Will Possible?" In *Essays on Action and Events: Philosophical Essays*, vol. 1., essay 2. Oxford: Oxford, 2001.
- Haidt, Jonathan. *The Happiness Hypothesis: Finding Modern Truth in Ancient Wisdom*. New York: Basic Books, 2006.
- Johnson, Robert N. "Weakness Incorporated." *History of Philosophy Quarterly* 15, no. 3 (1998): 349-367.
- Kant, Immanuel. *Critique of Practical Reason* (1788). In *Practical Philosophy*, translated and edited by Mary J. Gregor. Cambridge: Cambridge University Press, 1999.

207

- Kant, Immanuel. *Groundwork of the Metaphysics of Morals* (1785). In *Practical Philosophy*, translated and edited by Mary J. Gregor. Cambridge: Cambridge University Press, 1999. Kant, Immanuel. *The Metaphysics of Morals* (1797). In *Practical Philosophy*, translated and edited by Mary J. Gregor. Cambridge: Cambridge University Press, 1999. Kant, Immanuel. *Religion*

*Within the Boundaries of Mere Reason* (1793). In *Religion and Rational Theology*, translated and edited by Allen W. Wood and George Di Giovanni. Cambridge: Cambridge University Press, 2001.

Korsgaard, Christine M. *The Sources of Normativity*. Cambridge: Cambridge University Press, 1996.

Korsgaard, Christine M. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press, 2009.

Korsgaard, Christine M. "From Duty and For the Sake of the Noble: Kant and Aristotle on Morally Good Action." In *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, ch. 6. Oxford: Oxford University Press, 2014.

Korsgaard, Christine M. *Fellow Creatures: Our Obligations to the Other Animals*. Oxford: Oxford University Press, 2018.

Morrison, Iain. "On Kantian Maxims: A Reconciliation of the Incorporation Thesis and Weakness of the Will." *History of Philosophy Quarterly* 22, no. 1 (2005): 73-89. Moss, Jessica. "Akrasia and Perceptual Illusion." *Archiv für Geschichte der Philosophie* 91 (2009): 119-156.

Moss, Jessica. *Aristotle on the Apparent Good: Perception, Phantasia, Thought, and Desire*. Oxford: Oxford University Press, 2012.

Muchnik, Pablo. *Kant's Theory of Evil: An Essay on the Dangers of Self-Love and the*

*Aprioricity of History* (Lanham, MD: Lexington Books, 2009).

Schapiro, Tamar. *Feeling Like It: A Theory of Inclination and Will*. Oxford: Oxford University Press, 2021.

Vujošević, Marijana. "Kant's Account of Moral Weakness." *European Journal of Philosophy* 27, no. 1 (2019): 40-54.

<sup>1</sup> This paper is an attempt to state more clearly the view I put forth in *Feeling Like It: A Theory of Inclination and Will* (Oxford: Oxford University Press, 2021), ch. 6.

<sup>2</sup> I am trying to sidestep more fine-grained interpretive debates here.

<sup>3</sup> Donald Davidson, "How is Weakness of the Will Possible?" in *Essays on Action and Events: Philosophical Essays*, vol. 1, essay 2 (Oxford: Oxford University Press, 2001). <sup>4</sup> "Suppose someone asserts of his lustful inclination that, when the desired object and the opportunity are present, it is quite irresistible to him; ask him whether, if a gallows were erected in front of the house where he finds this opportunity and he would be hanged on it immediately after gratifying his lust, he would not then control his inclination. One need not conjecture very long what he would reply. But ask him whether, if his prince demanded, on pain of the same immediate execution, that he give false testimony against an honorable man whom the prince would like to destroy under a plausible pretext, he would consider it possible to overcome his love of life, however great it may be. He would perhaps not venture to assert whether he would do it or not, but he must admit without hesitation that it would be possible for him." Immanuel

<sup>5</sup> The name comes from Henry Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge, 1990), I.2.II.

<sup>6</sup> Immanuel Kant, *Religion Within the Boundaries of Mere Reason*, in *Religion and Rational Theology*, trans. and ed. Allen W. Wood and George Di Giovanni (Cambridge: Cambridge, 2001), 6:24.

<sup>7</sup> For a detailed account of how I am carving out the concept of inclination, along with an account of the philosophical method I am employing, see my *Feeling Like It*, ch. 1. <sup>8</sup> The literature on how to reconcile Kant's Incorporation Thesis with his account of volitional weakness includes Marcia Baron, "Freedom, Frailty, and Impurity," *Inquiry* 36 (1993): 431-441; Robert N. Johnson, "Weakness Incorporated," *History of Philosophy Quarterly* 15 (1998): 349-367; Iain Morriison, "On Kantian Maxims: A Reconciliation of the Incorporation Thesis and Weakness of the Will," *History of Philosophy Quarterly* 22 (2005): 73-89; Pablo Muchnik, *Kant's Theory of Evil: An Essay on the Dangers of Self-love and the Aprioricity of History* (Lanham, MD: Lexington Books, 2009); and Marijana Vujošević, "Kant's Account of Moral Weakness," *European Journal of Philosophy* 27 (2019): 40-54.

<sup>9</sup> Christine Korsgaard mentions this as a gap in his theory in her "From Duty and For the Sake of the Noble," in *The Constitution of Agency: Essays on Practical Reason and Moral Psychology* (Oxford: Oxford University Press, 2014). There she does not try to fill the gap, but in later work she seems to endorse a certain view, at least as a feature of her brand of Kantianism, if not as a reading of Kant. The view is that inclinations are perceptions of reasons, in some non-realist sense. Korsgaard, *Self-Constitution: Agency, Identity, and Integrity* (Oxford: Oxford University

addressing here.

<sup>10</sup> Immanuel Kant, *Groundwork of the Metaphysics of Morals*, in Gregor, *Practical Philosophy*, 4:405.

<sup>11</sup> Immanuel Kant, *Metaphysics of Morals*, in Gregor, *Practical Philosophy*, 6:409.

<sup>12</sup> Kant, *Critique of Practical Reason*, 5:74.

<sup>13</sup> Proponents of this type of view include Dennis Stampe, T. M. Scanlon, Talbot Brewer, Sergio Tenenbaum, and others. See my discussion of the perceptual analogy in *Feeling Like It*, ch. 3. <sup>14</sup> Jessica Moss uses this idea to interpret Aristotle on *Akrasia*. “*Akrasia* and Perceptual Illusion,” *Archiv für Geschichte der Philosophie* 91 (2009): 119-156 and *Aristotle on the Apparent Good: Perception, Phantasia, Thought, and Desire* (Oxford: Oxford University Press, 2012). I am making no claim about how to read Aristotle.

<sup>15</sup> See Hilary Bok’s detailed defense of practical standpoint compatibilism in her book, *Freedom and Responsibility* (Princeton, NJ: Princeton University Press, 1998).

<sup>16</sup> Putting it this way is a little ambiguous. Does the condition of having an inclination generate the task in the first place, or do we face the task independently, and then inclinations suggest possible ways of meeting it? I have not tried to argue the point, but I believe I am committed to the former. I am grateful to Kyla Ebels-Duggan for pressing me on this.

<sup>17</sup> Schapiro, *Feeling Like It*, ch. 4.

<sup>18</sup> Kant comes closest to saying this when he refers to the “pathological self” striving to make claims as if it were your whole self. *Critique of Practical Reason* 5:74. But Kant is not consistent in the way he writes about having an inclination.

<sup>19</sup> Christine M. Korsgaard, *Fellow Creatures: Our Obligations to the Other Animals* (Oxford: Oxford University Press, 2018), 3.2.

<sup>20</sup> I am eliding many qualifications and details here. I do want to take on Korsgaard's view that the instinctive mind is intelligent and capable of learning through experience. I do not want to take on her specific way of spelling out the difference between the instinctive mind and the rational mind, as laid out in *Fellow Creatures*, 3.2.4. I think the way to arrive at an account of the rational mind is to start by trying to characterize our distinctive relation to the instinctive part of our own minds, which is what I try to do in *Feeling Like It*.

<sup>21</sup> Or the elephant, as Jonathan Haidt argues in *The Happiness Hypothesis: Finding Modern Truth in Ancient Wisdom* (New York: Basic Books, 2006).

<sup>22</sup> Korsgaard relies on the Platonic analogy between the parts of the soul and the parts of a constitutionally-ordered polity. *Self-Constitution*, ch. 7. I explain the shortcomings of this analogy in *Feeling Like It*, ch. 5.

<sup>23</sup> The moral law is a special case of an incentive, one that does not conform to the model of an incentive of inclination. Yet another sort of incentive is one I will identify in section VIII. <sup>24</sup>

This is a point that I should have developed more clearly in *Feeling Like It*, chs. 5 and 6. <sup>25</sup>

Christine M. Korsgaard, *Sources of Normativity* (Cambridge: Cambridge University Press, 1996), "The Authority of Reflection," and elsewhere.

<sup>26</sup> I am making no claims that this representation has to be "conscious," in the sense that it is readily available to you as an explicit thought. That said, the theory as a whole is addressed to you, insofar as you are trying to figure out what it takes to act well in relation to your inclinations

Schapiro, *Feeling Like It*, ch. 1. As such, it is addressed to an agent who is interested in governing herself properly. To spell out what that takes, I have to individuate steps in a

212

procedure that need not always be forefront in our minds. Self-governed action is necessarily self-conscious, in that it necessarily involves some awareness of our own agency. But that does not imply that we are necessarily paying attention to that awareness.

<sup>27</sup> See the references in note 8 for discussions of the relevant texts.

<sup>28</sup> Kant, *Groundwork* 4:400.

<sup>29</sup> See also Schapiro, *Feeling Like It*, ch. 6.

<sup>30</sup> This is arguably another point of agreement with Kant.

<sup>31</sup> Actually, on my view, weakness does not need to involve acting against your better judgment, or the stronger reason. The core of weakness is dealing badly with your inclinations. You could escape into your animality and still in some sense end up acting in a way that might be described as being in accordance with your better judgment. But the identification of what you are doing with what you would be doing if you acted on your better judgment would be strained, since what you are doing would be guided by a defective maxim, whereas what you would be doing would be guided by a proper maxim.

<sup>32</sup> On Korsgaard's mature view, efficacy presupposes autonomy. Korsgaard, *Self-Constitution*. I do not mean to be denying that here. Although you may be efficacious in the pursuit of another cocktail, you are also fleeing your nature at this moment, and in that sense, there is a real question as to *who* is efficacious in that pursuit.

<sup>33</sup> Davidson, "How is Weakness of the Will Possible?"

